DOCUMENT RESUME

ED 246 851 IR 011 180

AUTHOR Borchardt, Donald A.
TITLE A Computer Assisted System for Grading Papers.
PUB DATE [80]
NOTE 13p.
PUB TYPE Reports - Descriptive (141) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Software; Evaluation Criteria; Evaluation Methods; *Grading; Higher Education; *Peer Evaluation; Problem Solving; *Student Evaluation; Theater Arts; Worksheets; *Writing (Composition); *Writing Evaluation

ABSTRACT

A computer assisted system called PEERRATE was used in a theater history course at Rutgers, the State University of New Jersey in Newark, in fall 1980, in order to give students an opportunity to anonymously grade one another's written work and to compare their own ratings with those of fellow students and the instructor. The PEERRATE system is a fairly sophisticated program, written in FORTRAN, that mathematically eliminates rater biases and prints customized rating reports for individual students. It was used to complement a problem-solving class procedure in which students study aspects of theatre history by working in groups and discussing open-ended problems. PEERRATE was used to parallel a teaching strategy--guided design--in which group cooperation is emphasized. It was also used for grading and comparing grades for student reports. The program automatically produces a measure of each student's skill in rating. Part of the rating project involved discussion and agreement on criteria to use in writing the reports. The response to PEERRATE in theater history, psychology, English, composition, and graduate business courses has been generally positive. Using the system in theatre history was particularly helpful for giving students perspective on their own writing. Five references are listed. Tables show rating scales used, criteria definitions, a typical customized printout for an individual student, and achievement scores, rating qualities, and rater abilities in ascending order. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED246851

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A COMPUTER ASSISTED SYSTEM FOR GRADING PAPERS

by

Donald A. Borchardt
Associate Professor
Theatre Arts and Speech
Rutgers, The State University of New Jersey
Newark, New Jersey

IR011180

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Donald A. Borchardt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Introduction

A computer assisted system called PEERRATE developed by Professor Purnell Benson at Rutgers, The State University of New Jersey in Newark, was used in a Theatre History course, Fall 1980, in order to give students an opportunity to anonymously grade one another's written work and to compare their own ratings with fellow students and the instructor. This system was used to complement a problem-solving class procedure in which students study aspects of theatre history by working in groups and discussing open-ended problems. The work was accomplished in a step-by-step process in which the problem was defined, information was gathered, possible solutions were considered, constraints were discussed, a choice was made. This was followed by an analysis, evaluation and final recommendation for solving the problem. Students wrote papers in which they described the work they had done and explained the solution they found. The purpose of this paper is to explain how the computer was used for grading and comparing grades for student reports.

PEERRATE can aid instructors of large classes when term papers are assigned. Instead of the instructor reading all of the papers, students read subsets of papers, and with the accumulated scores for the entire class the rating system mathematically calculates the grade for each paper and shows component ratings as well. The computer was used in this instance, however, to parallel a teaching strategy, Guided Design, in which group cooperation is emphasized. By reading and grading one another's papers the students continued to be involved in a decision-

making process by discussing and establishing criteria. Then, the rating procedure was objectified by using a numerical rating and by minimizing rater bias mathematically. The system can be used to grade written essays, oral reports, team reviews, or co-written tests.

The PEERRATE System

The PEERRATE system is a fairly sophisticated program written in Fortran that mathematically eliminates rater biases and prints customized rating reports for individual students. In this case each paper was given five ratings, three were anonymous peer ratings from class members, one was a self rating, and the fifth was the instructor rating. By viewing the final printout a student could see his overall score in addition to ratings on selected components of the paper which were agreed on as criteria.

Once the groups completed their work, the required summary papers were submitted in four copies. Papers were given identification numbers instead of names and then collated to insure proper distribution; that is, they were spread out so that no one set of papers could be lumped together. Students were reissued the reports of their (now anonymous) peers. Grading was done in each of the five categories comprising the criteria. Inherent in any rating task where the participants rate only a subset of the total group is the possibility of individual rater bias (whether too lenient or too stiff) distorting the ratings. Also, there is the human problem of student tendency to award classmates

with an "A" and, consequently, using only the high end of the rating scale. The PEERRATE computer program solves these procedural problems by taking student-provided ratings as input, reducing rater bias, adjusting the rating distribution to a pre-set mean and deviation, and then printing a report for each participant that includes not only his calculated grade but also an estimate of his efficiency as a grader.[1]

Even if an entire class of raters tends to use only the upper end of a rating scale, the intervals between performers (or in this case, papers) remain. It is a simple algebraic maneuver for the computer to spread out the part of the scale that has been used, maintaining the rank orders between papers but setting a new mean and deviation. This mean and deviation can be set up by an outside authority (usually the instructor) who knows what the group as a whole has produced.

The PEERRATE program automatically produces a measure of each student's skill in rating. It calculates four separate but related measures of rating consistency. First, the individual grades given by each rater are compared with the grades given to the same performers by the class as a whole. Second, a general "bias" score is calculated for each rater. Third, the range of the rating scale that the rater uses compared with the range used by all other raters is computed as a "difference" score. Fourth, after adjusting for systematic bias and for the range used, the computer prints the standard deviation of the residual error that remains between the individual rater's ratings and the adjusted class ratings. If students also rate their own work, the discrepancy between self-rating and class

rating can also be computed. All these measures of rating skill are combined by the computer into a single composite score for rating ability. This composite score could conceivably serve as one component in a student's course grade. Further information on the details of the PEERRATE program and it's statistical validity can be found in the papers by Benson.[2,3,4] Procedural information for using PEERRATE with term papers is detailed in a paper by Lutkus.[5]

The Criteria

Part of the rating project involved discussion and agreement on criteria to use in writing the reports. Classes vary in the criteria they feel are important but in this case it was decided that there would be five categories. Although each might have been assigned different weights, students thought each category was equally important, so each was to be considered 20% of the total. TABLE I shows the rating scale and defines the criteria. Students could describe the work done by their own group including the decisions made and the method used for making the decision. They could also explain what their own contributions had been. For example, one might have consulted outside sources that others in the group had not, or, a particular insight into the problem and finding a solution might be explained as another portion of the report. The various steps used to solve the problem were included in a third category and the time spent on each may have varied depending on the composition of the group and how they worked together. Sources of information was the fourth category and here each report might explain the most useful or helpful references and how they were incorporated

TABLE I: RATING SHEET SHOWING RATING SCALE AND DEFINED CRITERIA

PEERRATE

Theatre History NAME_____

Your Paper I.D. #_____ Date_____

EACH STUDENT WILL READ THREE PAPERS IN ADDITION TO HIS/HER OWN AND RATE THEM ACCORDING TO THE ESTABLISHED CRITERIA. EACH CRITERION SHOULD BE RATED. DO NOT TRY TO ADJUST A LOW RATING IN ONE CRITERION WITH A HIGH IN ANOTHER TO BALANCE THE OVERALL TOTAL.

Each criterion should be rated on a 1 to 20 scale. Select individual numbers from the rating scale (1 through 20) to indicate your perception of the quality of work.

RATING SCALE

1--2--3--4	5--6--7--8	9--10--11--12	13--14--15--16	17--18--19--20
Terrible	Poor	Adequate	Good	Excellent

CRITERIA DEFINED:

Group Evaluation: This is the work done by the group such as discusssion and decision-making and how well it is reported. It will be weighted 20% of the total.

Individual Contribution: This is the work contributed by the writer which may or may not agree with the group decisions. It will be weighted 20% of the total.

Problem-Solving Technique: This refers to the various steps taken toward solving the problem such as identifying the problem, gathering information, constraints, possible solutions, etc. It will be weighted 20% of the total.

Sources: This refers to where the individual or group obtained information e.g. personal experience, primary and secondary sources, films, pictures, etc. It will be weighted 20% of the total.

Facts of the Problem: This refers to Instructions and Gathering Information. It will be weighted 20% of the total.

Paper's I.D. Number	Group Eval.	Indiv. Contrib.	Prob-Solv.	Sources	Facts of Prob.	

into the decision-making process. The fifth criterion, facts of the problem, included handout material given in class as well as information they would have gathered in their research.

In the grid (TABLE I) below the "defined criteria," students could rate each category for three papers and include a self rating score as well. No overall score was given. When all of the papers were rated, they were collected and collated. Scores from the instructor and from student raters were used as data for computer input.

Customized Printout

TABLE II shows a typical customized printout for an individual student. In the upper right hand corner is the overall score for performance on the assignment. For this student, Identification Number 11, the score is 81. Directly below is an item rating adjusted to norm for each of the categories included as criteria. This may show a student which areas were strongest in his paper and which were weakest, and, if the same criteria were used again, where more time might be spent trying to improve the score.

The next columns of numbers show the ratings from each rater including the self rating (I.D. Number 11) and the instructor rating which is Identification Number 1. The instructor rated all categories lower than the student rated himself, the closest relationship in scores being in the categories of "group evaluation" and "individual contribution." Rater number 18 was a tough grader and gave lower scores than the instructor in all but one category, "problem-solving techniques." This may show how students understand terminology

TABLE II: CUSTOMIZED PRINTOUT

MACNEILL ID. NO. 11 PROJECT 311 NUMBER 1 ***PEER-RATING***

THEATRE HISTORY REPORT FALL 1980

**** OVERALL SCORE FOR PERFORMANCE OF ASSIGNMENT(S) 81
THIS IS SCORE RECEIVED ON ASSIGNMENT WITH 5 RATER RECORDS

POINTS FOR ITEMS ADJUSTED TO NORMS FOR GROUP:

ITEM NAME	POINT BASIS	ITEM RATING ADJ TO NORM	WEIGHT FOR ITEM
GROUP EVALUATION	100	87	0.200
INDIVIDUAL CONTRIBUTION	100	75	0.200
PROBLEM SOLVING TECHNIQUES	100	83	0.200
SOURCES	100	83	0.200
FACTS OF THE PROBLEM	100	79	0.200

OVERALL AND ITEM SELF-RATINGS BY PEER ON POINT BASIS ABOVE:
92 95 85 100 88 93

***REPORT OF RATINGS RECEIVED FROM INDIVIDUAL RATERS ON POINT BASIS ABOVE

RATER ID NO.	WT.	COMPARED WITH ADJ. GROUP RTG	SCORE GIVEN	ITEM RATING GIVEN BY RATER 1	2	3	4	5
1	1	RTGS FROM RTR:	84	90	80	83	83	83
5	1	RTGS FROM RTR:	99	100	95	100	100	100
7	1	RTGS FROM RTR:	83	85	68	85	93	85
11	1	RTGS FROM RTR:	92	95	85	100	88	93
18	1	RTGS FROM RTR:	76	88	60	88	78	65

BEST COPY AVAILABLE

differently from the instructor.

Comparative Rating Scores

TABLE III shows achievement scores, rating qualities, and rater abilities in ascending order. The first column shows that the performance score of 81 for the paper with I.D. Number 11 falls below the mean score of 85. There were only four other papers with lower scores in a class of 17 students. Under "Rating Quality" we can see that the quality of the score for I.D. Number 11 was 36 or about average for the class. This means that the raters for the paper were not all in perfect agreement. A high quality rating would be one in which all the raters seem to agree on scores. A low quality rating would be one in which raters had widely divergent scores in each of the categories.

The columns under "Rater Ability" show I.D. Number 11 to be a below average (for the class) rater. That is, the score of 83 indicates that there is enough difference from most other students in the class to suggest a different understanding of grading standards.

Conclusions

The response to PEERRATE in Theatre History, Psychology, English Composition, and Graduate Business courses has been generally quite positive. Data from a follow-up questionnaire completed by over 200 students in four different courses using PEERRATE indicated that 64% felt they learned more from the course (8% said "less," 37% expressed "little difference") than they would have without PEERRATE. Seventy-eight percent of the students felt that the

TABLE III: COMPARATIVE RATING SCORES

26A. ACHV SCORES, RATING QUALITIES, AND RATER ABILITIES IN ASCENDING ORDER

PERF SCORE			RATING QUALITY			RATER ABILITY		
ID NO.	SUB NO.	SCORE	ID NO.	SUB NO.	SCORE	ID NO.	SUB NO.	SCORE
1	1	*****	1	1	*****	9	9	70
5	5	70	9	9	70	16	16	70
9	9	75	5	5	74	17	17	76
4	4	76	4	4	76	12	12	82
10	10	78	16	16	80	11	11	83
11	11	81	10	10	82	3	3	84
7	7	83	3	3	83	10	10	84
6	6	83	15	15	84	4	4	84
12	12	85	13	13	84	5	5	87
2	2	87	11	11	86	18	18	87
16	16	87	8	8	86	14	14	88
14	14	89	2	2	90	15	15	90
17	17	89	14	14	90	2	2	90
15	15	90	5	6	91	7	7	91
18	18	90	17	17	91	6	6	92
8	8	92	12	12	92	8	8	93
13	13	93	7	7	92	13	13	94
3	3	96	18	18	95	1	1	97

BEST COPY AVAILABLE

system was either definitely worth carrying over to other courses or at least worth further trial. Only 21% opposed the system after experiencing it.

Using PEERRATE in a Theatre History course was particularly helpful for giving students perspective on their own writing. First of all, when they rate papers written by peers they often see what good writers include and what poor writers omit. And assigning numbers on a rating scale takes some of the subjective mystery out of grading. Second, it was noted that students avoided pseudo intellectualism knowing their peers as well as the instructor would be reading their papers. Finally, to see the scores given by peers in comparison to those given by the instructor suggests that grading will very likely differ depending on who is doing the reading. Although the teacher gives the final grade it is interesting to see that fellow students and teachers rate higher or lower than the writer himself thinks he deserves.

REFERENCES

1. Borchardt, Donald A. and Lutkus, Anthony D. "Peerrated Responses to Problem Solving in Theatre History." Proceedings. National Conference on Guided Design. University of West Virginia. Morgantown, West Virginia, 1980.

2. Benson, Purnell H. "A Computerized System of Student Grading of Student Assignments." Third Annual New Jersey Conference on the Use of Computers in Higher Education. New Brunswick, New Jersey: New Jersey Educational Computer Network, Inc., 1976, 21-30.

3. Benson, Purnell H. A Supervised Student Grading System. American Statistical Association, Social Statistics Section, 1977.

4. Benson, Purnell H. Documentation for the Use of PEERRATE. Rutgers University Center for Computing and Information Services, Newark, New Jersey 07102.

5. Lutkus, Anthony D. The Use of PEERRATE: A Computerized System of Student Term Paper Grading. Paper presented at the 86th annual convention of the American Psychological Association, Toronto, Canada, August 1978. (ERIC # ED 167887).